Predicting metabolic adaptation
from networks of mutational paths

Christos Josephides and Peter S. Swain*

SynthSys—Synthetic and Systems Biology, School of Biological Sciences, University of Edinburgh
Mayfield Road, Edinburgh EH9 3BF, UK

Competition for substrates is a ubiquitous selection pressure faced by microbes, yet intracellular trade-offs can prevent cells from metabolizing every type of available substrate. Adaptive evolution is constrained by these trade-offs, but their consequences for the repeatability and predictability of evolution are unclear. Here we develop an eco-evolutionary model with a metabolic trade-off to generate networks of mutational paths in microbial communities and show that these networks have descriptive and predictive information about the evolving communities. We find that long-term outcomes, including community collapse, diversity, and cycling, have characteristic evolutionary dynamics that determine the entropy, or repeatability, of mutational paths. Although reliable prediction of evolutionary outcomes from environmental conditions is difficult, graph-theoretic properties of the mutational networks enable accurate prediction even from incomplete observations. In conclusion, we present a novel methodology for analyzing adaptive evolution and report that the dynamics of adaptation are a key variable for predictive success.

Elucidating the factors that influence the emergence, diversity, and stability of microbial communities is a central interest in both ecology and evolution [1]. To predict and control community structure and function, it is necessary to understand how interactions between microbes and the environment manifest as selective pressures driving microbial adaptation and diversification [2].

Microbes frequently grow on mixtures of metabolic resources where competition for growth-limiting substrates is a ubiquitous selection pressure. Surveys have revealed that microbes do not simultaneously use all available substrates; instead, each species in a community specializes to a few substrates [3, 4]. This observation is anticipated by ecological (resource-ratio) theory, which posits that constraints in the capacity of an organism to use multiple substrates are necessary for diversifying selection in homogeneous environments [5], and is supported by competition experiments where trade-offs in using one substrate over another maintain metabolic polymorphisms [6, 7, 8, 9, 10].

The conditions under which a microbial population can invade another, and whether a stable community can be formed, are the subject of ecological invasion analysis. These analyses typically either assume competition between infinitesimally-varying phenotypes [11, 12] or are not concerned with the mutational paths [13, 14] that may be generated through successive mutations and invasions. Advances in experimental evolution, however, now enable tracking of microbial lineages for hundreds of generations [15, 16] and expose how sequences of mutations shape the evolutionary process through competition between possibly disparate phenotypes [8]. Moreover, these experiments reveal evolutionary trajectories with both parallel and unique dynamics [17, 18], as well as variability in long-term outcomes [10].

To investigate how the interplay between metabolic constraints, environmental conditions, and the distribution of mutations influences the adaptation process, we developed a model that combines microbial chemostat ecology with an evolutionary process. At the ecological level, microbes compete for two growth-limiting substrates, and we incorporate a phenomenological metabolic trade-off between the consumption of one substrate relative to the other. Cells inherit this phenotype through nearly-faithful clonal reproduction, but rare mutations can change the degree of substrate specialization. We do not restrict the size-effect of mutations so that, in the extreme case, any phenotype can mutate to any other. We determine and analyze all mutational paths and stable microbial communities that may arise to describe the entire adaptation process.

We begin by introducing the eco-evolutionary model and show that adaptation dynamics over mutational paths can be visualized as a network through constructing a Markov process over a finite set of possible communities. Multiple long-term behaviors emerge, including quasi-periodic outcomes and outcomes with many exclusive stationary states. Next, we show that these evolutionary outcomes are not restricted to

51 distinct environments and that small environmental perturbations can change one type of outcome to another.
52 In spite of this sensitivity to environmental conditions, the processes that lead to each evolutionary outcome
53 have characteristic adaptation dynamics, which are reflected in graph-theoretic properties of the mutational
54 paths. Finally, we show that evolutionary outcomes can be predicted using these network properties from
55 incomplete observations of evolving microbial communities — even without knowledge of environmental
56 conditions.
57

Results

Microbial ecology in the chemostat with two substrates

60 Microbial evolution is often studied in chemostats, and we focus on a chemostat model of microbial ecol-
61 ogy [19, 20]. To investigate metabolic adaptation on multiple substrates, we extend the standard chemostat
62 model [21] to include two growth-limiting substrates, which we name u and v (Supplementary Section 1.1).
63 The two substrates are individually sufficient for microbial growth — i.e. they are perfectly substitutable [22].
64 Both substrates are added continuously to the chemostat, and cells grow in proportion to their concentra-
65 tions. The chemostat is diluted at a constant rate, and populations must therefore grow at least as fast as
66 the dilution rate to survive being washed out. At steady-state, the growth rate of a surviving population is
67 equal to the chemostat dilution rate (Fig. 1A).
68   Cells in the chemostat encounter both u and v substrates but cannot simultaneously specialize to using
69 both. To investigate how a metabolic constraint affects adaptation [23], we consider a phenomenological
70 trade-off between the import capacity for the two substrates (Supplementary Section 1.2). We parameterize
71 the metabolic specialization of each cell by a number between zero and one, s: values near zero indicate that
72 the cell specializes to v; values near one indicate u specialization. Cells inherit their phenotype (their value
73 of s) from their parent nearly always without variation through clonal replication and cells cannot change
74 their phenotype (Fig. 1B).
75   We can interpret the metabolic specialization, s, in two ways, using a common mathematical formulation
76 (Supplementary Section 1.5). In the first interpretation, cells can always import both substrates but increase
77 the rate of import for u at the expense of decreasing the import rate for v. The maximal import rate for
78 u is multiplied by s; the maximal import rate for substrate v is multiplied by $(1 - s)$ (Supplementary
79 Section 1.3). Such a constraint may arise if a finite resource is shared between the production of the u and v
80 permeases [24, 25, 26] or through antagonistic pleiotropy [27, 6, 7]. In the second interpretation, cells adopt
81 a mixed strategy (following evolutionary game theory [28]) of randomly switching between two metabolic
82 programmes, each exclusive to only one substrate. Then, the probability of adopting the u metabolic
83 programme is s and the probability of adopting the v programme is $(1 - s)$ (Supplementary Section 1.4).
84   We model the microbial life cycle through a sequence of substrate import, metabolism, and growth
85 processes. Specifically, we structure the microbial population to discrete growth states that the cell progresses
86 through to replicate. Cells metabolize u and v molecules at substrate-specific rates and obtain a substrate-
87 specific yield in terms of growth state increments. Cells replicate when they have accumulated enough yield
88 to exceed a threshold number of growth states (Fig. 1C).
89   A cell will almost always produce offspring that inherit its metabolic phenotype, but, rarely, a mutation will
90 result in a phenotypically-distinct population. The new mutant will compete with the pre-existing resident
91 population for extracellular substrates. We follow the theory of adaptive dynamics and include phenotype
92 mutations only on evolutionary timescales [11, 12]: we assume that mutations are sufficiently rare that
93 mutant phenotypes emerge only after the chemostat model has reached steady state. The ecological and
94 evolutionary timescales are separated in this weak-mutation limit.
95   In our spatially homogeneous model at most two populations can co-exist on two substrates because of
96 the competitive exclusion principle [29]. A mutant population is at first rare and must grow faster than
97 the dilution rate to survive wash-out in the chemostat. When the chemostat contains a single phenotypic
98 population, for example, invasion by a mutant has three possible outcomes: the mutant becomes extinct,
99 the resident becomes extinct, or the two phenotypes co-exist in a community (Fig. 1D).

Markov process of mutation-limited adaptation

101 Frequency-dependent selection gives rise to multiple mutational paths, which are contingent on the sequence
102 and outcome of mutation and invasion events. Rare mutations generate a series of competition events
103 between the resident and mutant populations with different metabolic phenotypes. Successful invasion
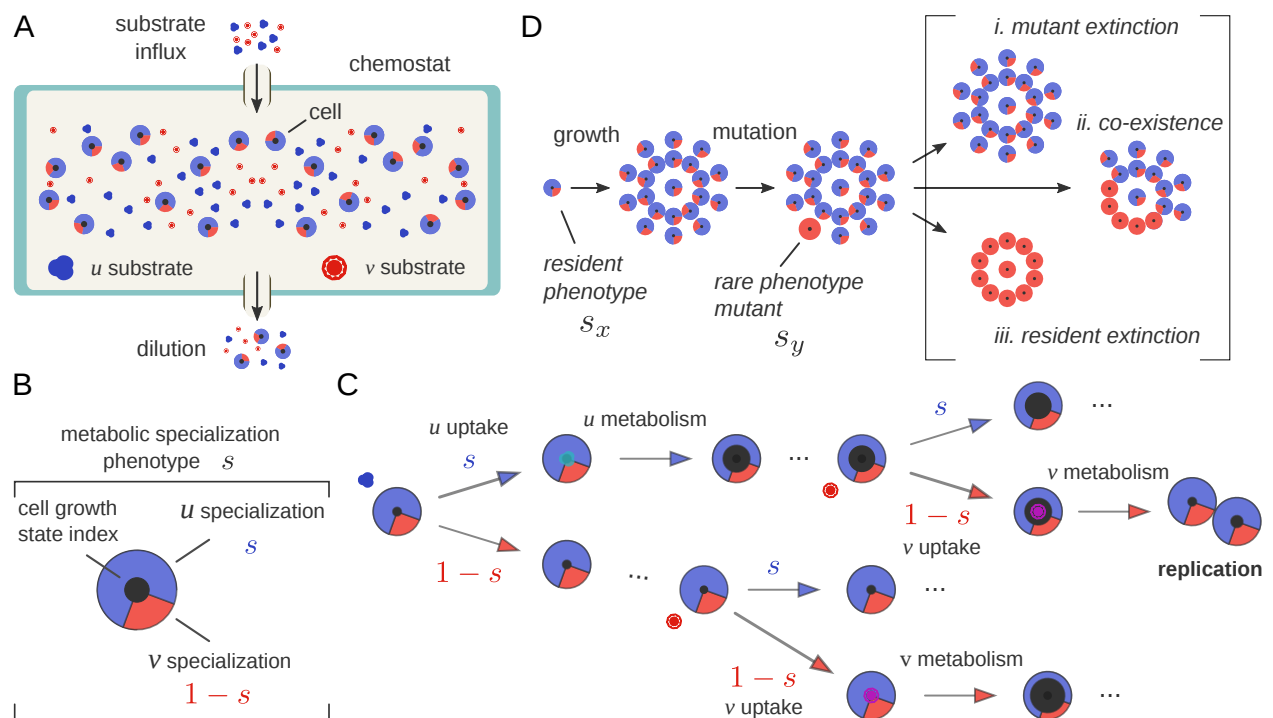

Figure 1: We model microbial ecology in the chemostat with two growth-limiting, substitutable, substrates (u and v) and a metabolic trade-off where specialization in one substrate reduces specialization in the other. **A.** The chemostat is a homogeneous environment with continuous addition of the two substrates and continuous dilution at fixed rates. The model is parameterized by the chemostat's dilution rate, and a pair of influx rates, import rates, metabolic rates, and yields for the two substrates. **B.** The metabolic specialization of a cell is parameterized by a number between zero and one, s: values near zero indicate a v-specialist; values near one indicate u-specialist; intermediate values are substrate generalists. **C.** Cells increment their growth state by different amounts depending on the yield of the substrate. Cells replicate when their growth state exceeds a critical value. When a cell encounters a u substrate molecule, the probability that it will successful import this molecule is its phenotype value (s); when a cell encounters a v substrate molecule, the probability that a cell imports a molecule of v substrate is $(1-s)$. **D.** Competition between a (high abundance) resident phenotype and a (low abundance) invading phenotype follows a rare mutation event and has three possible qualitative outcomes.

104 by a new population modifies the chemostat environment through changing the steady-state levels of the
105 substrates, and therefore the context in which future mutation and invasion events occur is modified through
106 such ecological niche construction and destruction [23, 8, 2, 30, 18].
107 To generate the mutational paths, we first create an invasion map for the outcome of competitions between
108 dynamically-stable communities of resident phenotypes and possible mutant phenotypes that may invade.
109 To do so efficiently, we developed a dynamic programming algorithm to simulate invasion events assuming
110 rare mutations on a discretized phenotype space (Fig. 2A). Briefly, the algorithm treats the invasion map as a
111 tree: nodes are communities of phenotypes that can co-exist at steady-state and these are connected by edges
112 representing single mutation and invasion events. The algorithm constructs the tree by iteratively perturbing
113 each steady-state (parent node) to introduce a small mutant population. The resulting competition is
114 simulated (Supplementary Section 1.6), and the steady-state outcome is analyzed and recorded as a connected
115 steady-state (child node). We avoid redundant simulations, and so avoid recursive expansions of the tree,
116 by appropriately terminating branches (Supplementary Section 2.1).
117 To define a Markov process of adaptation, we require the probability that a community will transition
118 from one state to another as a result of a single mutation and invasion event (Supplementary Section 2.2).
119 The probability of such a transition is proportional to the abundance of cells in the source state and to
120 the propensity with which each cell generates the mutation that effects the transition to the target state
121 (according to the deterministic invasion map). We will only be concerned with the sequence of transitions
122 between one community state to another and not with the waiting time between such events. In this case,
123 the distribution of mutations is the most important factor for determining which of the possible states will
124 next follow a current state.
125 We use a uniform distribution of mutations in phenotype space centered on a resident phenotype, s_x,
126 with maximum mutation size ΔS_{max}. Mathematically, a mutant phenotype s_y can be generated by a cell
127 with phenotype s_x if $|s_y - s_x| \leq \Delta S_{max}$. The mutation probabilities are chosen so that all of the discrete s_y
128 values that can be generated from s_x are equally likely after correcting for boundary effects (Fig. 2B,D). The

3

129 distribution of mutations affects both the mutational paths and the stationary behavior of the adaptation
130 process, and we parse the invasion map using uniform distributions with different maximum mutation sizes
131 to investigate this dependency.

132   To complete the description of mutation-limited adaptation as a Markov process, we must choose an
133 initial distribution of ancestral states. Most laboratory evolution experiments start with isogenic strains,
134 and, similarly in our model, adaptation starts with equal probability from any viable population comprising
135 a single phenotype (rather than a community).

136   To determine the long-term evolutionary outcome of an adaptation process, we classify the states of the
137 Markov process as either transient or recurrent. Transient states represent microbial communities that may
138 only be visited once on a mutational path. A recurrent state, however, will always re-emerge once established,
139 and the endpoints of mutational paths are always recurrent states (Fig. 2C,E). Recurrence occurs either when
140 a microbial community cannot be invaded by any mutant that can be generated from the community (an
141 evolutionarily-stable state [28]) or if there is a sequence of mutation and invasion events that returns to the
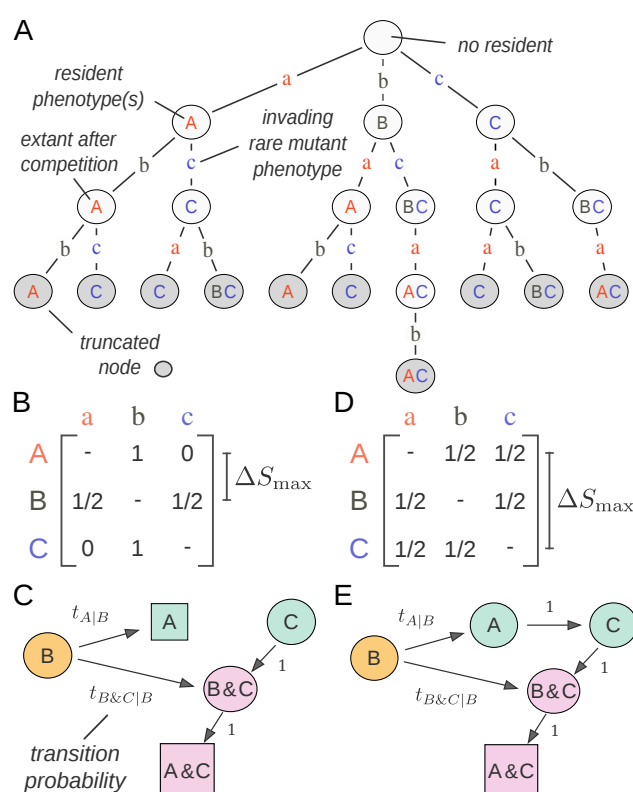142 community.

Figure 2: The Markov process of adaptation is generated from two components — simulations of ecological invasion and calculations of the transition probabilities based on the probabilities of each potential mutation — and is demonstrated with an example that has three phenotypes (labeled A, B, and C). **A.** To discover all dynamically-stable populations and communities, we simulate mutation and invasion events in a space of discrete phenotypes assuming rare mutations. We construct an invasion map as a tree: nodes are steady-states labeled with extant phenotypes. Each state is perturbed with a small mutant population and the resulting steady-state after competition is connected to the parent node. The algorithm dynamically truncates the tree to avoid any recursion. **B.** To calculate transition probabilities between resident states, we parse the invasion map in A with a mutational process. We assume uniformly distributed mutations with a maximum mutation size, $\Delta S_{\max}$, centered on the resident phenotype. The matrix elements show the conditional probability that phenotype s_x generates mutant phenotype s_y. **C.** The Markov process in B visualized through its directed network graph. States are classified as either transient (circles) or recurrent (squares). **D-E.** As in B-C, but with a larger maximum mutation size ($\Delta S_{\max}$). The recurrent states, as well as the mutational paths of the process, depend on the distribution of mutations.

Hierarchy of evolutionary outcomes

144 To investigate the effect of the environment and mutational parameters on the adaptation process, we ran-
145 domly sampled $10,000$ environmental parameter sets (Supplementary Section 3.1). For each set of parame-
146 ters, we calculated the invasion map using 11 discrete phenotype values ranging from a pure v-specialist to a
147 pure u-specialist: $s \in \{0.0, 0.1, \ldots, 0.9, 1.0\}$. We parsed each invasion map with uniform mutation distribu-
148 tions with increasing maximum mutation sizes, $\Delta S_{\max} \in \{0.1, 0.2, \ldots, 0.9, 1.0\}$, to generate ten adaptation

149 processes. We then analyzed the 100,000 resulting adaptation processes to determine their long-term evo-
150 lutionary behavior.

151   We classify the evolutionary outcome of an adaptation process according to the number and type of its
152 recurrent states (Fig. 3A). At a high level, we differentiate between processes with either a single recurrent
153 state or multiple recurrent states. A recurrent state may be a dimorphic community, with two-coexisting
154 phenotypes, or a monomorphic population with a single phenotype. We classify phenotypes as either (u or
155 v) substrate specialists, when $s \in \{0.0, 1.0\}$, or substrate generalists, when $s \in \{0.1, 0.2, \ldots 0.8, 0.9\}$, and we
156 distinguish between the possible community combinations of specialists and generalists in a recurrent state.

157   We observed two qualitatively different behaviors in processes with multiple recurrent states, which we
158 named 'multi-stationary' and 'cycling'. Multi-stationary outcomes have many exclusive recurrent states
159 where each recurrent state is reached with some probability. We found that the number of recurrent states
160 and their stationary probability depended on the maximum mutation size (Fig. 3B, C). In these cases,
161 we expect replicate evolution experiments to eventually show divergent phenotypic variability [8, 10, 31].
162 Cycling outcomes, on the other hand, are processes where no microbial community is completely resistant to
163 invasion, and the adaptation process continually cycles between a set of states through mutation and invasion
164 events. This cycling can be either periodic (Fig. 3D *left*) or aperiodic and unpredictable — albeit confined
165 to a closed class of recurrent states (Fig. 3D *right*). These community changes occur at the evolutionary
166 time-scale [32]. For example, in a dimorphic community of one specialist and one generalist population,
167 mutation and invasion events can drive the generalist to extinction; this extinction leaves an unexploited
168 metabolic niche that a newly-emergent generalist population can fill, thereby repeating the cycle (Fig.3D).
169 Both of these evolutionary outcomes are qualitatively reproduced in stochastic simulations of mutational
170 paths on a continuous phenotype space (Supplementary Section 2.3).

171   Visualizing the networks not only conveys information on the evolutionary outcomes but also on the
172 adaptation dynamics. For example, we found that some adaptation processes had bottle-necked mutational
173 paths: all mutational paths have to transition through a few common community states before reaching
174 a recurrent state (Fig. 3E). We will later quantify these network properties to characterize the adaptation
175 dynamics and construct a predictive model.

Sensitivity of adaptation to environmental conditions

177 In light of the multiplicity and complexity of evolutionary outcomes, we investigated how particular outcomes
178 are associated with the environmental parameters of the model. These parameters describe the chemostat
179 set-up (substrate influx and dilution rates) and the properties of the two substrates (their yields, maximal
180 import rates, and metabolic rates).

181   Although some evolutionary outcomes were more likely to be associated with certain environments, no
182 clear pattern emerged (Fig. 4A). Using a nearest-neighbors algorithm to estimate the density of evolutionary
183 outcomes, we observed that metabolically diverse (dimorphic) communities are most likely to emerge when
184 the influx rates for the two substrates are similar and when the dilution rate is low. As the disparity in supply
185 of the substrates increases, we find that communities undergo evolutionary cycling, perhaps signalling the
186 onset of ecological collapse [33]. At even greater ratios of substrate influx, microbial communities disappear,
187 leaving a population of a single specialist as the only evolutionary outcome.

188   Even by using combinations of environmental parameters to characterize outcomes, substantial unpre-
189 dictability remained. We developed a hierarchical classification model, which combines six statistical estima-
190 tors to make predictions (Supplementary Section 3.2). Its overall performance (a mean recall of 0.78 over the
191 eight evolutionary outcomes), however, did not suggest that a reliable predictive map from environmental
192 parameters to evolutionary outcomes could easily be formed.

193   We can partly understand this unpredictability by considering how the different evolutionary outcomes
194 are distributed in parameter space, where distinct clusters do not appear to exist. For each outcome, we first
195 performed hierarchical clustering using the pairwise Euclidean distance in a standardized parameter space
196 (Supplementary Section 3.3) and then aligned these clusters with samples from other outcomes to assess
197 their overlap (Fig. 4B). In general, while parameter sets that produced the same outcomes did form clusters,
198 these were not distinct: a sample or cluster from one outcome is typically proximal to samples or clusters
199 from other outcomes. Environmental perturbations are therefore often likely to cause a qualitative shift in
200 the long-term community.

201   In particular, monomorphic specialists 'permeate' parameter space, and are typically close to all other
202 outcomes. By calculating the distribution of shortest-distances from one type of outcome to either the same
203 outcome or one of the other seven possibles outcomes, we can estimate the magnitude of the environmental
204 change required to both maintain the type of outcome and to alter one outcome to another (Supplementary
205 Section 3.4). Assuming that the probability of environmental perturbation is inversely proportional to its
206 magnitude, a change in the environment is approximately as likely to preserve a diverse community as to
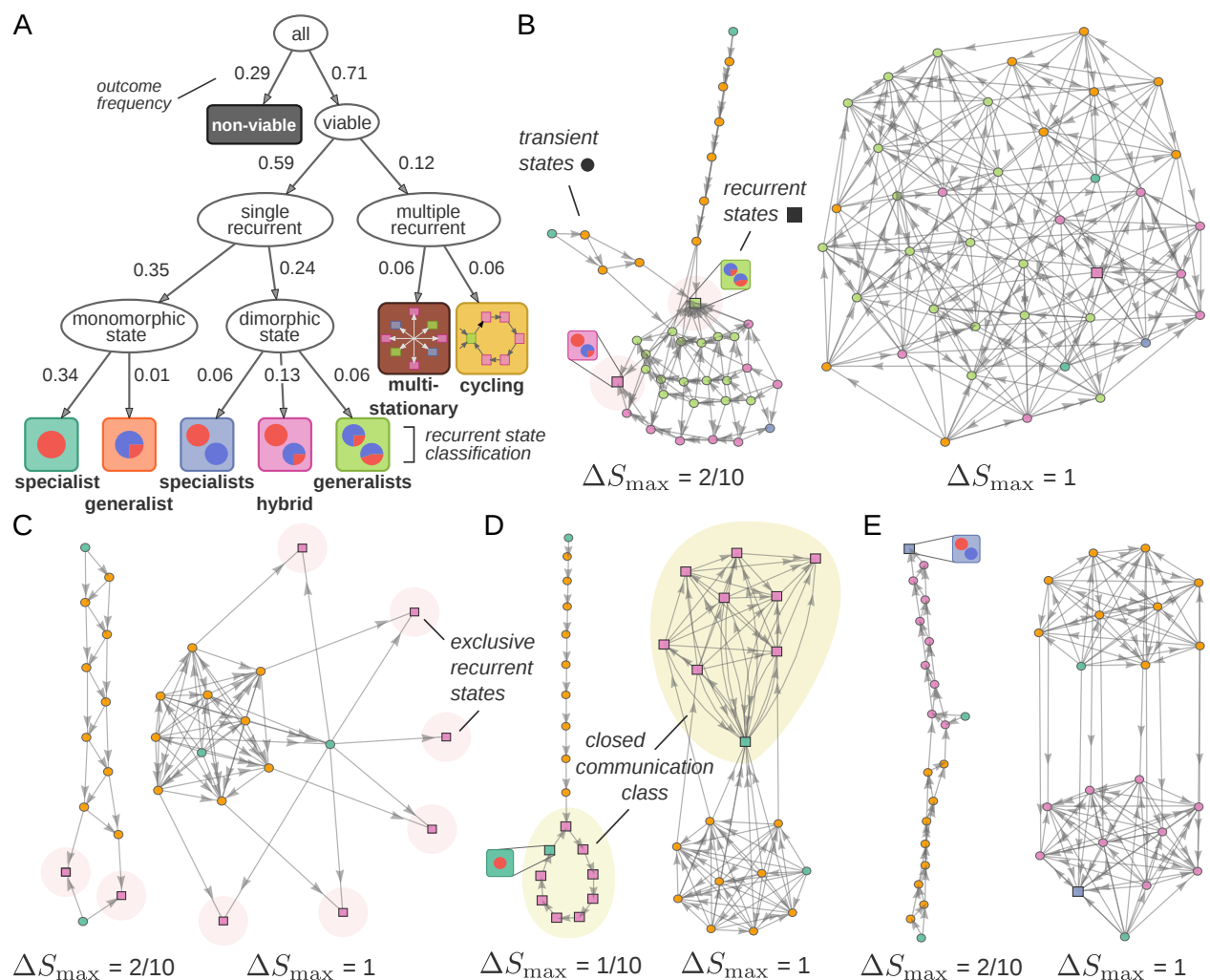
Figure 3: The model generates multiple evolutionary outcomes and complex networks of mutational paths. Qualitative network properties of adaptation dynamics can be visualized and conveyed through graphical representation. **A.** We hierarchically classify adaptation processes according to their evolutionary outcome using the number and type of recurrent states. Labels show the outcome frequencies obtained from parameter sampling. **B-E.** Networks of mutational paths from four sets of environmental parameters demonstrate the scope of the model's dynamic and stationary behavior. Circles are transient states; squares are recurrent states; colors denote the number and type of metabolic phenotypes in each (microbial community) state following A. **B.** The maximum mutation size ($\Delta S_{\max}$) affects the mutational paths and long-term outcomes of adaptation (compare recurrent states between *left* and *right* networks). **C.** A process with multiple, non-connected, recurrent states has more than one evolutionarily stable state, each of which is reached with some (stationary) probability. **D.** A process where multiple recurrent states are connected exhibits quasi-periodic evolutionary cycling. **E.** An example of a potentially bottle-necked process. The network consists of two highly-connected (top and bottom) subgraphs, which are themselves connected via only a few mutation and invasion transitions.

207 lead to its collapse to a monomorphic specialist (Fig. 4C, *right*). The converse, however, is not true: a
208 monomorphic specialist population is far less likely to transition to a diverse community (Fig. 4C, *left*).
209 We conclude that prediction of evolutionary outcomes from environmental conditions, even in simple
210 chemostat environments, is likely to be challenging. Evolutionary outcomes are only the endpoints of the
211 adaptive process, however, and we next investigate if the dynamics of these processes (the mutational paths)
212 are characteristic of their endpoints.

Adaptation dynamics and mutational paths

214 The mutational paths of the Markov process describe the dynamics of adaptation, and the processes we
215 construct have tens to millions of mutational paths in spite of a relatively small number (eleven) of discrete
216 phenotypes and a maximum of two co-existing populations in each community.
217 We analyze adaptation dynamics by enumerating the mutational paths of a process (Supplementary
218 Section 4.1), calculating their probabilities, and determining their properties (Fig. 5A). For example, we
219 define the path length as the number of state transitions, from an ancestral state to a recurrent state, in one
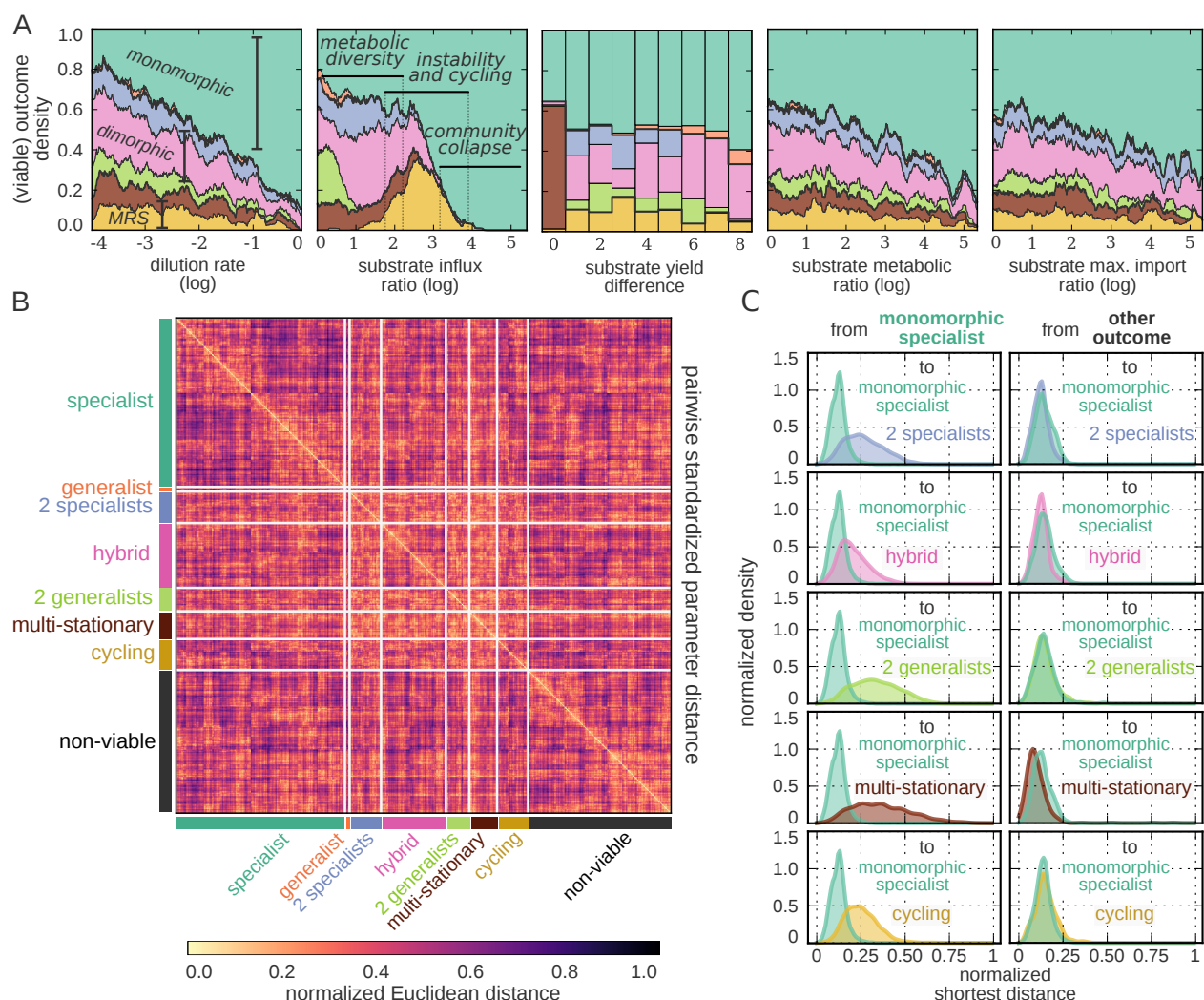
Figure 4: Complex association between the environment and evolutionary outcomes undermines the stability of microbial communities and complicates prediction. **A.** Evolutionary outcomes are more likely to be associated with some parameter values, but without a simple dependence. We estimated the empirical outcome densities from parameter sampling using a nearest-neighbors algorithm. For all parameter combinations except the dilution rate, we show densities at absolute value because of the symmetry about zero. **B.** Evolutionary outcomes form clusters in parameter space but these are not distinct. We used Euclidean distance in a standardized parameter space to perform within-outcome hierarchical clustering and then aligned clusters with samples from other outcomes to assess their overlap. We only plot a representative sample (1%) of the dataset, which preserves the observed frequencies of each outcome. **C.** Environmental perturbations often cause community collapse and only rarely create diverse communities. We calculated the standardized parameter distance from each process to the closest process in all other outcome classes to form the distribution of shortest-distances. Monomorphic specialists are generally closer to other monomorphic specialists than they are to outcomes with diverse communities (*left*). In contrast, processes with diverse outcomes are as close to each other as they are to monomorphic specialists (*right*). Here 'other outcome' refers to the labeled outcome in each panel that is not a monomorphic specialist.

220 realization of the adaptation process (a single run of an evolution experiment).

221 The degree to which adaptive evolution is repeatable is of long-standing interest [34, 35], but analyses
222 are typically restricted to models with static fitness landscapes [36]. To quantify the repeatability of adap-
223 tation in our model, where fitness landscapes change dynamically through eco-evolutionary feedbacks, we
224 mathematically define repeatability as the entropy of the path distribution. If replicate experiments are
225 likely to follow only a few mutational paths, the entropy is small and repeatability is high; if each replicate
226 experiment follows a different mutational path, the entropy is large and repeatability is low.

227 In our model, the repeatability of adaptation dynamics is associated with the evolutionary outcome of the
228 adaptation process and varies with the maximum mutation size (Fig. 5B). Path entropy typically increases
229 with maximum mutation size as more mutational paths become possible. Notably, however, dimorphic
230 specialist outcomes have a non-monotone path entropy with maximum mutation size (Fig. 5B, blue line)
231 because a few mutational paths with large mutations come to dominate the process. We find that processes
232 where monomorphic specialist populations are evolutionarily stable have the most repeatable dynamics; in

233 contrast, processes with metabolically diverse communities have the least repeatable adaptation dynamics.

234 To compare the dynamics between adaptation processes, we constructed a condensed dynamical profile
235 for each through calculating six statistical properties of its mutational paths. These properties were: the
236 number of paths, the mean and variance of geodesic (shortest) path length, the mean and variance of all
237 paths, and the mean minimum cut size (the smallest number of edges that must be removed from the graph
238 to disconnect an initial state from a recurrent state — a measure of the extent of bottle-necks in the process).

239 We analyzed the six mutational path properties for processes with different evolutionary outcomes and at
240 different maximum mutation sizes (Fig. 5C). The effect of increasing the maximum mutation size depends
241 on the evolutionary outcome of the process, but the qualitative trends are consistent. Specifically, larger
242 mutations decrease both the mean and the variance of the geodesic path length because recurrent states
243 can be reached via fewer mutation and invasion events of larger effect. Larger mutations have a similarly
244 decreasing, albeit smaller, effect on the mean and variance of the overall path length. Finally, we found that
245 a larger maximum mutation size generally leads to processes with fewer bottle-necks (higher minimum cut
246 size), particularly for processes with a single recurrent state.

247 The distributions of mutational path properties are multi-modal (Fig. 5C), where the peaks correspond
248 to evolutionary outcomes low in the outcome hierarchy (Fig. 3A), suggesting that processes with different
249 evolutionary outcomes have characteristic adaptation dynamics. We used discriminant analysis to find
250 linear combinations of the six path properties that maximize the separation of the seven viable evolutionary
251 outcomes according to their dynamics. We found that different outcomes, particularly those higher in the
252 outcome hierarchy (Fig. 3A), occupy different regions in a two-dimensional projection of the discriminant
253 space (Fig. 5D). We can therefore conclude that mutational paths characteristically identify processes' long-
254 term outcomes, and we contrast this positive finding with the difficulty of obtaining an association between
255 environmental parameters and evolutionary outcomes (Fig. 4).

256 The path properties we calculate quantify statistical aspects of the adaptation process; however, enumera-
257 tion of paths becomes computationally prohibitive for larger networks and also relies on correct classification
258 of states. We next present an alternative graph-theoretic approach that relaxes these constraints and forms
259 the basis of a model for predicting evolutionary outcomes from observations of adaptation dynamics.

Predicting evolutionary outcomes from networks of mutational paths

261 To motivate our predictive model, we consider an evolution experiment as the progressive construction of a
262 network of mutational paths. In such an experiment, the resident microbial populations in a chemostat [37]
263 are periodically assayed, and the transitions between resident communities are used to construct the network.
264 To simulate such a procedure, we re-sampled the $100,000$ adaptation processes in our dataset to obtain
265 networks at varying stages of completion (Supplementary Section 5.1). Incomplete networks may contain
266 spurious recurrent states, which obfuscate the process's evolutionary outcome (Fig. 6A, *top*). These states
267 appear recurrent because the mutation and invasion events that lead away from these states have not yet
268 been observed in the experiment. The predictive objective, then, is to forecast the true recurrent states of
269 the complete network from an incomplete sample.

270 Our predictive model relies on quantifying how networks change with accumulating observations. As we
271 have shown earlier, evolutionary outcomes have characteristic mutational paths. To use the information en-
272 coded in the entire network of mutational paths at once, we measure graph-theoretic properties that quantify
273 abstract features of the network. Specifically, for each network, we calculated six centrality measures [38] that
274 characterize different aspects of network topology (Supplementary Section 4.2 and Supplementary Figure 6).
275 Though less interpretable from an ecological or evolutionary perspective compared to path properties (such
276 as path length) a network's centrality distributions nevertheless contain valuable predictive information. For
277 each network we calculate the six centrality distributions and summarize each through its mean and variance.
278 As mutation and invasion events are observed, new vertices and edges are added to the network, and the
279 resulting changes in network topology are reflected in changes in the centrality statistics (Fig. 6A, *bottom*).

280 We identified both invariant and informative topological changes as networks are progressively completed
281 (Supplementary Section 5.2). We found that some centrality statistics converge to their complete-network
282 value even when the network is mostly incomplete — remaining invariant as more vertices and edges are
283 added — while others change considerably during network construction (Fig. 6B). Nevertheless, we noted
284 that the most informative centrality statistics (those with distinct distributions for the different outcomes)
285 converged gradually (Supplementary Figure 7).

286 Differences in network topology, even for incomplete networks, were sufficient to enable reliable prediction
287 of evolutionary outcomes using a statistical-learning approach. Centrality statistics provide a projection
288 of the mutational path network, through its topology, to a low-dimensional space. We trained a classifier
289 to predict evolutionary outcomes from these centrality features using data from incomplete and complete
290 networks, the degree of network completion, and the maximum mutation size (Supplementary Section 5.3).
291 To avoid biasing against minority outcomes, we used the unweighted mean $F1$ score (over the outcome
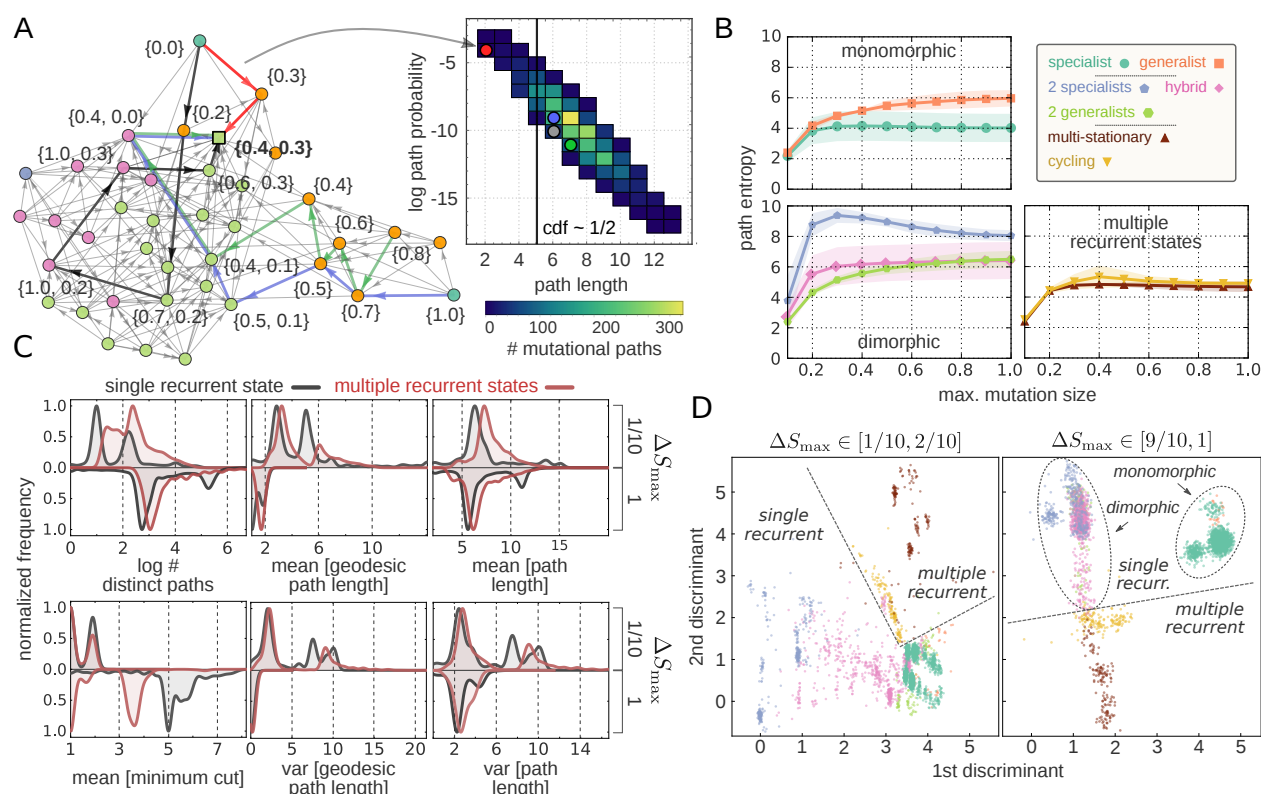
8

Figure 5: Processes with different evolutionary outcomes have characteristic adaptation dynamics, which can be quantified and compared through the properties of their mutational paths. **A.** We enumerate mutational paths to generate the path distribution. Paths begin at a monomorphic state (with equal probability) and end at a recurrent state. The path probability decreases in proportion to the length of the path. Most mutational paths in a network are of intermediate length, and only a few paths are either short or long. Statistical properties of adaptation, like the mean and variance of the path length, are calculated from the path distribution. **B.** The repeatability of adaptation dynamics depends on the long-term evolutionary outcome and typically increases with maximum mutation size. The entropy of the path distribution quantifies repeatability (high entropy implies low repeatability). Points show the mean path entropy for networks with different outcomes; shaded regions are ± the standard deviation. **C.** Mutational path statistics vary with the maximum mutation size and the process's evolutionary outcome. We plot the distributions for six statistical path properties, grouping by the number of recurrent states and two extremes of maximum mutation size. The distributions are multi-modal, and the peaks correspond to outcomes lower in the classification hierarchy (Fig. 3A). **D.** Processes with different evolutionary outcomes have characteristic adaptation dynamics. Linear combinations of the six path properties in C that best separate the adaptation processes in a two-dimensional projection were identified through linear discriminant analysis. Lines and ellipses were drawn manually.

classes) as the optimization metric. Neither the degree of network completion or the maximum mutation size will typically be known during an experiment, and we marginalized over both these variables during testing (Supplementary Figure 8). The performance of the classifier improves with increasing network completion and is best at intermediate maximum mutation sizes (Fig. 6C). At 50% network completion, the classifier predicts the correct long-term outcome for about 85% of all adaptation processes (average $F1$ score > 0.7), which rises to 98% at 60% graph completion ($F1 = 0.8$).

We have shown here that our predictive approach can be useful in forecasting changes in the community, including the loss of (metabolic) diversity, through incomplete observations of adaptation dynamics. Detecting the onset of community transitions is an enduring problem, but most approaches focus on catastrophic transitions and analyze the dynamics of a single dynamical model [39, 40, 33]. Nevertheless, the loss of biodiversity following environmental disturbance can, as we have shown, involve a series of ecological transitions mediated by multiple mutation and invasion events, and our network approach can address this challenge.
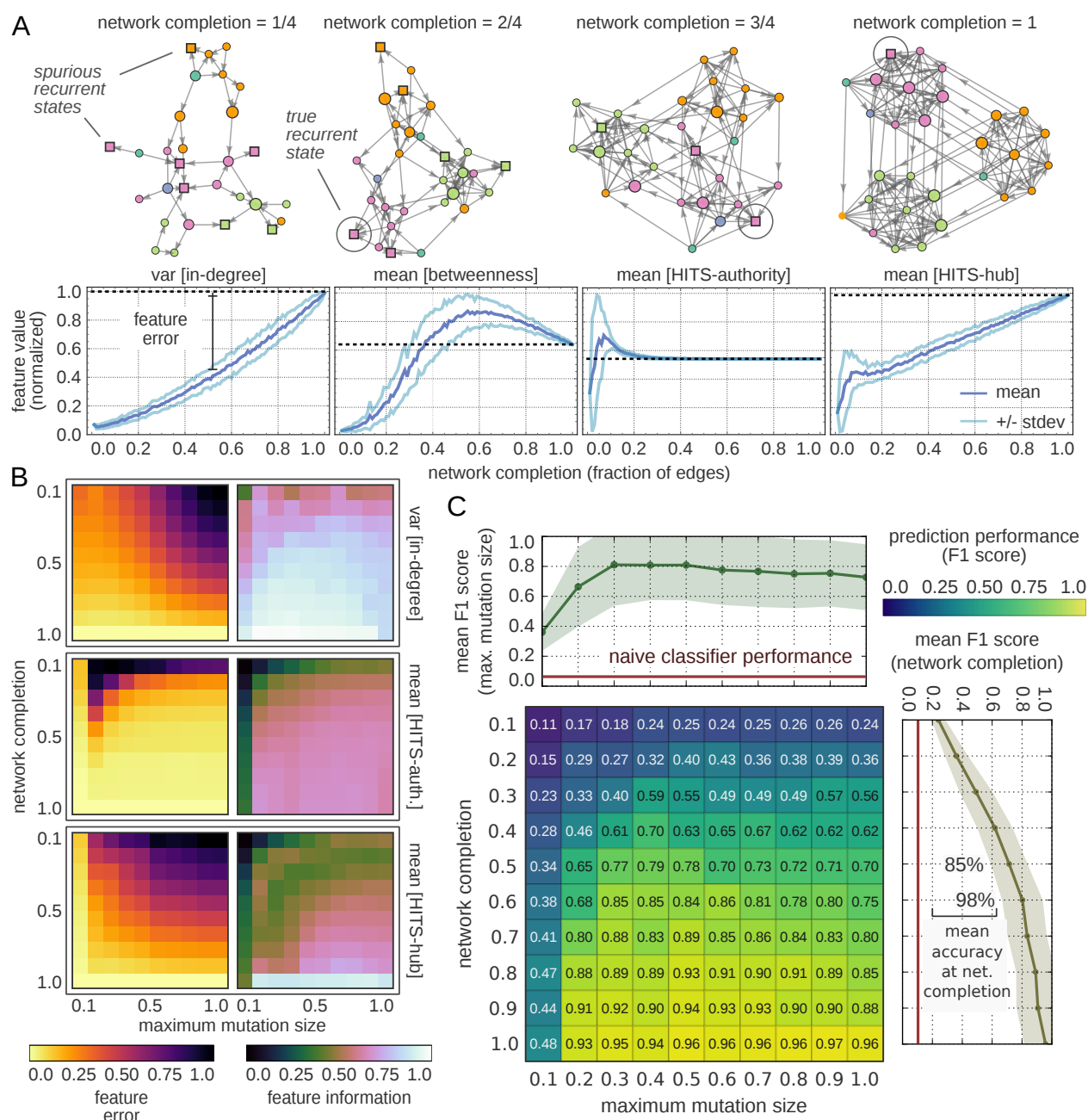
Figure 6: Evolutionary outcomes can be predicted from graph-theoretic properties of incomplete networks of mutational paths. To generate incomplete networks we sampled networks to varying degrees of completion. For both complete and incomplete networks, we calculated and summarized six network centralities, which characterize different aspects of network topology, through their mean and variance. Statistics for the twelve centralities summarize mutational networks through a low-dimensional space. **A.** The state of network completion affects its topological properties, which is reflected in the centrality statistics. For an example process (*top*), we show the progression of four centrality statistics (*bottom*) to highlight the variability in their convergence as a function of network completion. Dark blue lines show the mean and light blue lines show the standard deviation for 100 network samples at different stages of network completion. **B.** We characterized the centrality statistics both by how quickly they converge to their complete-network value (feature error, *left*) and by how well they can discriminate between evolutionary outcomes (feature information, *right*). The plots show the mean of feature error (the normalized difference between the complete- and incomplete- values of the centrality statistic) and the mean of the mutual information between the centrality statistic and the evolutionary outcomes. We found both invariant (e.g. mean of HITS authority) and informative (e.g. variance of in-degree) centrality statistics (Supplementary Section 5.2) as well as a dependency on the maximum mutation size. **C.** Evolutionary outcomes can be predicted with a statistical classifier trained on the twelve centrality statistics from partial and complete networks. We assessed classifier performance via the unweighted mean of the $F1$ score, taken over the seven evolutionary outcomes, and report the mean performance on the test set for a ten-fold cross-validation. *Top* and *right* line plots show mean performance, with standard deviation in shaded regions, taken over the columns (network completion) and rows (maximum mutation size). The red line is the benchmark performance of a naive classifier, which predicts following the empirical frequencies of the outcomes.

Discussion

We have shown that a simple model of metabolic adaptation comprising a trade-off in the use of one type of substrate over another can generate surprisingly rich behavior. On one hand, this behaviour is sufficiently complex that it is difficult to predict the outcomes of adaptation, even qualitatively, from environmental parameters. On the other, we have demonstrated that statistical properties of the mutational paths can uniquely characterize long-term outcomes. Thus, similarly to structural indicators in evolutionary game theory [41, 42] and statistical indicators in ecology [39, 40, 33], the dynamics of adaptation are a key variable for the successful prediction of community outcomes.

Our approach combines elements from the replicator-mutator equation [43, 44] and adaptive dynamics [11, 45, 12] to couple an ecological model with an evolutionary process. Ecological interactions between cells and the environment induce eco-evolutionary feedbacks [11, 45, 23, 2, 30, 18]: resident phenotypes define environmental niches for invasion by mutants, which in turn shape niches for future invasion. Such behavior is predicted to occur, for example, in the gastrointestinal tract [46, 47]. We formalized adaptation as a discrete-time Markov chain because the evolutionary model initiates ecological invasion via rare mutation events. This formulation emphasizes the dynamics of adaptation, i.e. the mutational paths [36, 34], rather than focusing on isolated phenotype competitions or evolutionary 'end-points'.

Several extensions can be made, but are likely to only increase the complexity of the dynamics observed. For example, the substrates can be made partly substitutable [22, 5] or their number increased. Growth could be parameterized by several phenotypic traits combining multiple metabolic [48, 49] or cellular [26, 50] trade-offs. We ignored stochastic extinction of rare mutants, and we anticipate that such effects will change the statistical properties, but not the long-term outcomes, of adaptation. Finally, the model operates in the 'strong-selection, weak-mutation' regime [51], therefore we omit both multiple simultaneously invading mutants (clonal interference [52, 53]) and the occurrence of multiple mutations occurring within one invading lineage [54, 55].

The model demonstrates the importance of trade-offs for generating metabolic complexity in microbial communities [5, 1]. If a cell is able to evolve its response to one substrate unconstrained by its response to the other, then only one evolutionary outcome is possible — a phenotype that maximally depletes both substrates — because all other phenotypes will be unable to grow sufficiently fast in the chemostat to survive dilution [56, 21]. Metabolic trade-offs, however, preclude this single optimal phenotype and, together with dynamically changing environmental niches [8, 2, 18] and a limited distribution of possibly large-effect mutations [57], generate complex adaptation dynamics and long-term behaviors. Intracellular trade-offs are expected to be common [48, 26, 50], and our results support the idea that the resulting frustration of optimal behaviors is a major factor generating the complexity observed in microbial communities.

Methods

Model simulations

We modified the chemostat model [21] to include two substrates and structured populations (Supplementary Section 1.5), and simulated the model to steady-state using the numerical integration libraries in Mathematica 10 [58] (Supplementary Section 1.5-1.6). Models were dynamically generated, by adding and removing invading and extinct populations, as required for the invasion map (Supplementary Section 2.1). The invasion map was parsed to calculate the state transition probabilities for the Markov process (Supplementary Section 2.2).

Data analysis

We analyzed the adaptation processes from $10,000$ samples of environmental parameters (Supplementary Section 3.1). The path distribution, statistics of path properties (Supplementary Section 4.1), and centralities (Supplementary Section 4.2) were calculated for each network and are available as a tabulated dataset. Data was analyzed using the Python `scikit-learn`[59] library for statistical models. Incomplete networks were generated by re-sampling the complete-network dataset (Supplementary Section 5.1). Mutual information was calculated using non-parametric entropy estimation [60] (Supplementary Section 5.2). The predictive model was based on a Random Forest classifier (Supplementary Section 5.3).

Acknowledgments

356 Trust PhD studentship; P.S.S. acknowledges support from the Scottish Universities Life Science Alliance.

357 Author contributions

358 C.J. and P.S.S. designed research; C.J. performed research and analyzed data; C.J. and P.S.S. wrote the
359 paper.

360 References

361 [1] Stefanie Widder, Rosalind J Allen, Thomas Pfeiffer, Thomas P Curtis, Carsten Wiuf, William T Sloan,
362 Otto X Cordero, Sam P Brown, Babak Momeni, Wenying Shou, et al. Challenges in microbial ecology:
363 building predictive understanding of community function and dynamics. *ISME J*, 10:2557–2568, 2016.

364 [2] David M Post and Eric P Palkovacs. Eco-evolutionary feedbacks in community and ecosystem ecology:
365 interactions between the ecological theatre and the evolutionary play. *Philos Trans R Soc Lond B Biol*
366 *Sci*, 364(1523):1629–1640, 2009.

367 [3] Dana E Hunt, Lawrence A David, Dirk Gevers, Sarah Preheim, Eric J Alm, and Martin F Polz.
368 Resource partitioning and sympatric differentiation among closely lated bacterioplankton. *Science*,
369 320(5879):1081–1085, 2008.

370 [4] Richard Baran, Eoin L Brodie, Jazmine Mayberry-Lewis, Eric mmel, Ulisses Nunes Da Rocha, Romy
371 Chakraborty, njamin P Bowen, Ulas Karaoz, Hinsby Cadillo-Quiroz, and Ferran d others Garcia-Pichel.
372 Exometabolite niche partitioning among sympatric soil bacteria. *Nat Commun*, 6:8289, 2015.

373 [5] David Tilman. Resources: a graphical-mechanistic approach to competition and predation. *Am Nat*,
374 116(3):362–393, 1980.

375 [6] Mark Lunzer, Arvind Natarajan, Daniel E Dykhuizen, and Antony M Dean. Enzyme kinetics, substi-
376 tutable resources and competition: from biochemistry to frequency-dependent selection in lac. *Genetics*,
377 162(1):485–499, 2002.

378 [7] Daniel E Dykhuizen and Antony M Dean. Evolution of specialists in an experimental microcosm.
379 *Genetics*, 167(4):2015–2026, 2004.

380 [8] Maren L Friesen, Gerda Saxer, Michael Travisano, and Michael Doebeli. Experimental evidence for sym-
381 patric ecological diversification due to frequency-dependent competition in escherichia coli. *Evolution*,
382 58(2):245–260, 2004.

383 [9] Shaobin Zhong, Stephen P Miller, Daniel E Dykhuizen, and Antony M Dean. Transcription, translation,
384 and the evolution of specialists and generalists. *Mol Biol Evol*, 26(12):2661–2678, 2009.

385 [10] David Healey, Kevin Axelrod, and Jeff Gore. Negative frequency-dependent interactions can underlie
386 phenotypic heterogeneity in a clonal microbial population. *Mol Syst Bio*, 12(8):877, 2016.

387 [11] Johan AJ Metz, Stefan AH Geritz, Géza Meszéna, Frans JA Jacobs, and Joost S Van Heerwaarden.
388 Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction. *IIASA*,
389 95(99), 1995.

390 [12] Fabio Dercole and Sergio Rinaldi. *Analysis of evolutionary processes: the adaptive dynamics approach*
391 *and its applications.* Princeton University Press, 2008.

392 [13] Daniel M Weinreich, Nigel F Delaney, Mark A DePristo, and Daniel L Hartl. Darwinian evolution can
393 follow only very few mutational paths to fitter proteins. *Science*, 312(5770):111–114, 2006.

394 [14] Barrett Steinberg and Marc Ostermeier. Environmental changes bridge evolutionary valleys. *Sci Adv*,
395 2(1), 2016.

396 [15] Christopher D Herring, Anu Raghunathan, Christiane Honisch, Trina Patel, M Kenyon Applebee,
397 Andrew R Joyce, Thomas J Albert, Frederick R Blattner, Dirk Van den Boom, Charles R Cantor,
398 et al. Comparative genome sequencing of escherichia coli allows observation of bacterial evolution on a
399 laboratory timescale. *Nat Genet*, 38(12):1406–1412, 2006.

400 [16] Sandeep Venkataram, Barbara Dunn, Yuping Li, Atish Agarwala, Jessica Chang, Emily R Ebel, Kerry
401 Geiler-Samerotte, Lucas Hérissant, Jamie R Blundell, Sasha F Levy, et al. Development of a com-
402 prehensive genotype-to-fitness map of adaptation-driving mutations in yeast. *Cell*, 166(6):1585–1596,
403 2016.

404 [17] Ludovic Pelosi, Lauriane Kühn, Dorian Guetta, Jérôme Garin, Johannes Geiselmann, Richard E Lenski,
405 and Dominique Schneider. Parallel changes in global protein profiles during long-term experimental
406 evolution in escherichia coli. *Genetics*, 173(4):1851–1869, 2006.

407 [18] Matthew D Herron and Michael Doebeli. Parallel evolutionary dynamics of adaptive diversification in
408 escherichia coli. *PLoS Biol*, 11(2):e1001490, 2013.

409 [19] Santiago F Elena and Richard E Lenski. Evolution experiments with microorganisms: the dynamics
410 and genetic bases of adaptation. *Nat Rev Genet*, 4(6):457–469, 2003.

[20] David Gresham and Maitreya J Dunham. The enduring utility of continuous culturing in experimental evolution. *Genomics*, 104(6):399–405, 2014.

[21] Hal L Smith and Paul Waltman. *The theory of the chemostat: dynamics of microbial competition*. Cambridge university press, 1995.

[22] Jesús Alberto León and Daniel B Tumpson. Competition between two species for two complementary or substitutable resources. *J Theor Biol*, 50(1):185–201, 1975.

[23] Ulf Dieckmann and Michael Doebeli. On the origin of species by sympatric speciation. *Nature*, 400(6742):354–357, 1999.

[24] Daniel M Stoebel, Antony M Dean, and Daniel E Dykhuizen. The cost of expression of escherichia coli lac operon proteins is in the process, not in the products. *Genetics*, 178(3):1653–1660, 2008.

[25] Gene-Wei Li, David Burkhardt, Carol Gross, and Jonathan S Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635, 2014.

[26] Andrea Y Weiße, Diego A Oyarzún, Vincent Danos, and Peter S Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc Natl Acad Sci*, 112(9):E1038–E1047, 2015.

[27] Daniel E Dykhuizen, Antony M Dean, and Daniel L Hartl. Metabolic flux and fitness. *Genetics*, 115(1):25–31, 1987.

[28] John Maynard Smith. *Evolution and the Theory of Games*. Cambridge university press, 1982.

[29] Garrett Hardin et al. The competitive exclusion principle. *Science*, 131(3409):1292–1297, 1960.

[30] Mickaël Le Gac, Jessica Plucain, Thomas Hindré, Richard E Lenski, and Dominique Schneider. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with escherichia coli. *Proc Natl Acad Sci*, 109(24):9487–9492, 2012.

[31] Meike T Wortel, Evert Bosdriesz, Bas Teusink, and Frank J Bruggeman. Evolutionary pressures on microbial metabolic strategies in the chemostat. *Sci Rep*, 6, 2016.

[32] Ulf Dieckmann, Paul Marrow, and Richard Law. Evolutionary cycling in predator-prey interactions: population dynamics and the red queen. *J Theor Biol*, 176(1):91–102, 1995.

[33] Lei Dai, Daan Vorselen, Kirill S Korolev, and Jeff Gore. Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science*, 336(6085):1175–1177, 2012.

[34] Alexander E Lobkovsky and Eugene V Koonin. Replaying the tape of life: quantification of the predictability of evolution. *Front Genet*, 3:246, 2012.

[35] J Arjan GM de Visser and Joachim Krug. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet*, 15(7):480–490, 2014.

[36] Alexander E Lobkovsky, Yuri I Wolf, and Eugene V Koonin. Predictability of evolutionary trajectories in fitness landscapes. *PLoS Comput Biol*, 7(12):e1002302, 2011.

[37] Aaron W Miller, Corrie Befort, Emily O Kerr, and Maitreya J Dunham. Design and use of multiplexed chemostat arrays. *J Vis Exp*, (72):e50262–e50262, 2013.

[38] Mark Newman. *Networks: an introduction*. Oxford university press, 2010.

[39] Marten Scheffer, Jordi Bascompte, William A Brock, Victor Brovkin, Stephen R Carpenter, Vasilis Dakos, Hermann Held, Egbert H Van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59, 2009.

[40] Stephen R Carpenter, Jonathan J Cole, Michael L Pace, Ryan Batt, WA Brock, Timmothy Cline, Jim Coloso, James R Hodgson, Jim F Kitchell, David A Seekell, et al. Early warnings of regime shifts: a whole-ecosystem experiment. *Science*, 332(6033):1079–1082, 2011.

[41] Matteo Cavaliere, Guoli Yang, Vincent Danos, and Vasilis Dakos. Detecting the collapse of cooperation in evolving networks. *Sci Rep*, 6, 2016.

[42] David A Gianetto and Babak Heydari. Sparse cliques trump scale-free networks in coordination and competition. *Sci Rep*, 6:21870, 2016.

[43] Karen M Page and Martin A Nowak. Unifying evolutionary dynamics. *J Theor Biol*, 219(1):93–98, 2002.

[44] Martin A Nowak and Karl Sigmund. Evolutionary dynamics of biological games. *Science*, 303(5659):793–799, 2004.

[45] Ulf Dieckmann and Richard Law. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J Math Biol*, 34(5-6):579–612, 1996.

[46] Elizabeth K Costello, Keaton Stagaman, Les Dethlefsen, Brendan JM Bohannan, and David A Relman. The application of ecological theory toward an understanding of the human microbiome. *Science*, 336(6086):1255–1262, 2012.

[47] Luke McNally and Sam P Brown. Building the microbiome in health and disease: niche construction and social conflict in bacteria. *Philos Trans R Soc Lond B Biol Sci*, 370(1675), 2015.

[48] I Gudelj, RE Beardmore, SS Arkin, and RC MacLean. Constraints on microbial metabolism drive evolutionary diversification in homogeneous environments. *J Evol Biol*, 20(5):1882–1889, 2007.

[49] Robert E Beardmore, Ivana Gudelj, David A Lipson, and Laurence D Hurst. Metabolic trade-offs and the maintenance of the fittest and the flattest. *Nature*, 472(7343):342–346, 2011.

13

[50] Thomas Ferenci. Trade-off mechanisms shaping the diversity of bacteria. *Trends microbiol*, 24(3):209–223, 2016.

[51] John H Gillespie. Molecular evolution over the mutational landscape. *Evolution*, 38(5):1116–1129, 1984.

[52] Philip J Gerrish and Richard E Lenski. The fate of competing beneficial mutations in an asexual population. *Genetica*, 102:127–144, 1998.

[53] H Allen Orr. The rate of adaptation in asexuals. *Genetics*, 155(2):961–968, 2000.

[54] Michael M Desai and Daniel S Fisher. Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics*, 176(3):1759–1798, 2007.

[55] Michael M Desai, Daniel S Fisher, and Andrew W Murray. The speed of evolution and maintenance of variation in asexual populations. *Curr biol*, 17(5):385–394, 2007.

[56] Mary M Ballyk and Gail SK Wolkowicz. Exploitative competition in the chemostat for two perfectly substitutable resources. *Math Biosci*, 118(2):127–180, 1993.

[57] Michael M Desai. Statistical questions in experimental evolution. *J Stat Mech*, 2013(01):P01003, 2013.

[58] Wolfram Research Inc. Mathematica 10, 2016.

[59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *J Mach Learn Res*, 12:2825–2830, 2011.

[60] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys Rev E*, 69(6):066138, 2004.